

CLAIMS

1. A method comprising:
receiving a document and an associated document type definition;
generating a mapping file from the document, comprising one or more
5 nodes, each node representative of a possible mapping of an
element of the document type definition to a portion of the
document;
generating one or more candidate paths from the mapping file, with each
candidate path representing a possible path from one node in the
10 mapping file to another node in the mapping file;
determining a score for each of the one or more candidate paths;
selecting one of the candidate paths based on the one or more scores; and
converting the one of the candidate paths into a language described by
the document type definition.
15
2. The method of claim 1, wherein the document is received after the
document type definition.
3. The method of claim 1, wherein the document type definition comprises
20 a Standard Generalized Markup Language and/or an Extensible Markup
Language document type definition.
4. The method of claim 1, wherein the determining a score for each one or
more candidate paths comprises:
25 determining a score based on compliance with the document type
definition without inferring additional tags for each of the one or
more candidate paths.
5. The method of claim 1, wherein the determining a score for each one or
30 more candidate paths comprises:

determining a score based on the document type definition with inferring tags for each of the one or more candidate paths.

6. The method of claim 1, wherein the determining a score for each one or more candidate paths comprises:

determining a score based on a recursive examination of each path for a predetermined extent from each node in the tree structure of the mapping file for each of the one or more candidate paths.

7. The method of claim 1, wherein determining one or more scores comprises determining two or more scores for each one of the one or more candidate paths and defining the highest of the two or more scores as the determined scored for the one of the one or more candidate paths.

8. The method of claim 1, wherein the document comprises a plurality of segments; and wherein the actions of generating one or more candidate, determining a score, selecting one of the candidate paths, and converting the one of the candidate paths are performed for each of the plurality of segments.

9. A method comprising:
receiving a document and an associated document type definition;
generating a mapping file from the document and the document type definition, with the mapping file comprising one or more nodes, each node representative of a possible mapping of an element of the document type definition to a portion of the document; and
disambiguating the mapping file based on the document type definition.

10. The method of claim 9, wherein the disambiguating comprises:

generating all of the permutations of the candidate paths from the mapping file, with each candidate path representing a possible path from one node in the mapping file to another node in the mapping file;

5 determining a score for each one or more candidate paths;
selecting one of the candidate paths based on the one or more scores; and
converting the one of the candidate paths into a language described by the document type definition.

10 11. A Standard Generalized Markup Language document-type-definition parseable file produced by a process comprising:
receiving a non-Standard Generalized Markup Language document;
receiving a Standard Generalized Markup Language document-type-
definition associated with the document;
15 disambiguating the document based on the Standard Generalized Markup Language document-type-definition, yielding disambiguated data;
and
converting the disambiguated data into a file parseable based on Standard Generalized Markup Language.

20 12. The Standard Generalized Markup Language document-type-definition parseable file produced by the process of claim 11, wherein receiving a Standard Generalized Markup Language document-type-definition associated with the document comprises receiving a Standard
25 Generalized Markup Language document-type-definition in the document.

13. A computer-readable magnetic, electronic, or optical medium comprising computer-executable instructions for:

- causing a computer to read a document and an associated document type definition;
 - causing a computer to generate a mapping file based on the document and the document type definition, with the mapping file comprising one or more nodes, each node representative of a possible mapping of an element of the document type definition to a portion of the document;
 - causing a computer to generate one or more candidate paths from the mapping file, with each candidate path representing a possible path from one node in the mapping file to another node in the mapping file;
 - causing a computer to determine one or more scores from the one or more candidate paths;
 - causing a computer to select one of the candidate paths based on the one or more scores; and
 - causing a computer to convert the one of the candidate paths into a language described by the document type definition.
14. A system comprising:
- means for receiving a document and an associated document type definition;
 - means for generating a mapping file based on the document and the document type definition, with the mapping file comprising one or more nodes, each node representative of a possible mapping of an element of the document type definition to a portion of the document;
 - means for determining one or more scores for one or more candidate paths, with each candidate path representing a possible path from one node in the mapping file to another node in the mapping file; and

means for selecting one of the candidate paths based on the one or more scores.

15. The system of claim 13, wherein the means for receiving, the means for
5 generating, the means for determining, and the means for selecting exist as
respective software modules in a memory coupled to one or more computer
processors or within various parts of a mainframe computer or within a SUN
Ultra 4000 Server or within an IBM-compatible personal computer.

10 16. A system for transacting in electronic commerce comprising:
a processor;
a storage device coupled to the processor;
software means operative on the processor for disambiguating
ambiguated data based on a document type definition.

15 17. The system of claim 16, wherein the software means comprises:
software means for generating all of the permutations of the candidate
paths from a mapping file of the ambiguated data, with each
candidate path representing a possible path from one node in the
20 mapping file to another node in the mapping file.

18. A computerized system comprising:
a document of ambiguous data;
a document type definition;
25 a mapper of ambiguous data, operatively coupled to the document and
operatively coupled to the document, yielding a mapping file
from the document and the document type definition;
a disambiguator operatively coupled to the mapping file and the
document type definition, yielding an output file; and
30 wherein the document type definition describes a markup syntax; and

wherein the output file complies with the syntax described by the document type definition.

19. The computerized system of claim 18, the system comprising:
5 a configuration file operatively coupled to the disambiguator which specifies predetermined settings and/or parameters of the disambiguator.
20. The computerized system of claim 18, the system comprising:
10 an activity log operatively coupled to the disambiguator that receives and record information that describes the activity of the conversion process of the disambiguator.
21. The computerized system of claim 18, the disambiguator comprising:
15 a permutater of one or more candidate paths from the mapping file, operatively coupled to the mapping file, with each candidate path representing a possible path from one node in the mapping file to another node in the mapping file;
a scorer of the one or more candidate paths, operatively coupled to the
20 permutater, yielding a corresponding number of one or more scores;
a selector of one or more candidate paths, based on the one or more scores, operatively coupled to the scorer, yielding a selected candidate path; and
25 a converter of the selected candidate path into a language described by the document type definition, operatively coupled to the selector of one or more candidate paths and operatively coupled to the document type definition.
- 30 22. The computerized system of claim 21, the disambiguator comprising:

a selector of segments, operatively coupled to the mapping file and
operatively coupled to the permutator, that receives the mapping
file and transmits a segment of the nodes of the mapping file to
the permutator;

5 a comparator operatively coupled to the mapping file and the permutator
that determines the extent of remaining segments in the mapping
file.

23. The computerized system of claim 21, the scorer comprising:
10 a tie-breaker operatively coupled to the permutator that selects one of a plurality
of candidate paths have equal scores.

24. A computer-readable magnetic, electronic, or optical medium
comprising:
15 a converter; and
a plurality of document type definitions associated with the converter.

25. A computer-readable magnetic, electronic, or optical medium
comprising:
20 a document;
a document type definition associated with the document; and
a converter operably coupled to the document and the document type
definition that converts the document into an output file that
complies with the document type definition.

25
26. A data structure stored on a computer-readable medium for representing
a possible mapping of an element of a document type definition to a portion of a
document comprising:
one or more segments.

30

27. The data structure of claim 26, wherein each of the one or more segments comprises:

- a field storing data representing a solid node;
- a field storing data representing a quantum node; and
- 5 a field storing data representing a terminal node.

28. The data structure of claim 27, wherein the data structure comprises two or more segments, and wherein two contiguous segments further comprising, a first segment and a second segment, are joined whereby the field storing data
10 representing a terminal node of the first segment further comprises the field storing data representing a solid node of the second segment.

29. A computer data signal embodied in a carrier wave and representing a sequence of instructions which, when executed by a processor, cause the
15 processor to perform:

- receiving a non-Standard Generalized Markup Language document;
- receiving a Standard Generalized Markup Language document-type-
definition associated with the document;
- disambiguating the document based on the Standard Generalized Markup
20 Language document-type-definition, yielding disambiguated data;
- and
- converting the disambiguated data into a file parseable based on Standard
Generalized Markup Language.

25 30. A computer data signal embodied in a digital data stream comprising data comprising:

- a representation of a solid node;
- a representation of a quantum node; and
- a representation of a terminal node;

30 wherein the computer data signal is generated by a method comprising:

generating a mapping file from the document and the document type definition, with the mapping file comprising one or more nodes, each node representative of a possible mapping of an element of the document type definition to a portion of the document.

5

31. A computer data signal embodied in a digital data stream comprising data comprising:

a representation of a Standard Generalized Markup Language document-type-definition parseable file;

10

wherein the computer data signal is generated by a method comprising: generating one or more candidate paths from a mapping file of an ambiguated document and a document type definition, with each candidate path representing a possible path from one node in the mapping file to another node in the mapping file;

15

determining a score for each of the one or more candidate paths; selecting one of the candidate paths based on the one or more scores; and converting the one of the candidate paths into Standard Generalized Markup Language document-type-definition parseable file described by the document type definition.

20

32. A method comprising:

providing a disambiguator;

disambiguating a first document based on a first DTD using the disambiguator;

25

disambiguating a second document based on a second DTD using the diasambiguator, with the second DTD being different than the first DTD.

33. A method of disambiguating electronic documents, comprising:

30

providing a set of two or more DTDs; receiving a first document for disambiguation; and

selecting at least one of the set of DTDs; and
disambiguating the first document based on the selected one of the set of
DTDs.

- 5 34. A method of disambiguating electronic documents, comprising:
receiving a first document for disambiguation, the document having first
and second portions, each portion being ambiguated;
disambiguating the first portion of the document and not the second
portion of the document; and
10 outputting a second document comprising the disambiguated first portion
of the document and second portion of the first document.

35. A method comprising selectively converting to one of a plurality of
markup languages.

15

20